DOCUMENT RESUME

ED 124 180 IR 003 572

AUTHOR	Pearson, Charls; Slamecka, Vladimir
TITLE	Semiotic Foundations of Information Science. Program Report No. 2.
INSTITUTION	Georgia Inst. of Tech., Atlanta. School of Information and Computer Science.
SPONS AGENCY	National Science Foundation, Washington, D.C. Office of Science Information Services.
PUB DATE	Mar 76
GRANT	NSF-OSI-GN-40952
NOTE	19p.; For the period February 1975-February 1976 ; for related document, see ED 107 229
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	Cognitive Processes; *Communication (Thought Transfer); Computer Oriented Programs; Higher Education; Information Processing; *Information Science; Information Theory; *Language; Language Research; Linguistics; Models; Semantics; *Semiotics; *Symbolic Language; Syntax; Word Recognition
IDENTIFIERS	Natural Language; Research Laboratories

ABSTRACT

This study is designed to help provide an understanding of the role semiosis plays in information processes. The object of study is the structure of various types of signs, and a determination of the relationship between sign structure and information properties. A review of the theory of sign structure is followed by an examination of syntactic levels of the universal structure model. Investigations into syntactic shape, semantic structure, and pragmatic structure are summarized. The semiotics laboratory of the Georgia Institute of Technology is described, and activities in semiotics and information science are reviewed. (CH)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED124180

Internal Research Memorandum

SEMIOTIC FOUNDATIONS OF INFORMATION SCIENCE

Progress Report No. 2

National Science Foundation Grant GN-40952

Charls Pearson
Vladimir Slamecka

Principal Investigators

March 1976

School of Information and Computer Science
Georgia Institute of Technology
Atlanta, Georgia 30332

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

IR 003572

ACKNOWLEDGMENTS

The work described in this progress report has been supported in part by Grant GN-40952 from the Office of Science Information, National Science Foundation. The project staff members responsible for this report are Albert N. Badre, assistant professor; Dan Canyas, graduate research assistant; Jeff Flowers, graduate research assistant; Charls Pearson, research associate; James D. Ray, Jr., graduate research assistant; Keum Shin, graduate research assistant; Phillip J. Siegmann, associate professor; Vladimir Slamecka, professor; and Pranas Zunde, professor.

The report covers the period of February 1975 to February 1976.

TABLE OF CONTENTS

		Page
	ACKNOWLEDGMENTS	i
	INTRODUCTION	1
I.	THEORY OF SIGN STRUCTURE	2
II.	SYNTACTIC DEVELOPMENTS OF THEORY	4
III.	INVESTIGATIONS INTO SYNTATIC SHAPE	5
IV.	INVESTIGATIONS INTO SEMANTIC AND PRAGMATIC STRUCTURE	7
V.	THE SEMIOTICS LAB	10
VI.	OTHER ACTIVITIES	11
	REFERENCES	13

INTRODUCTION

The purpose of the present study is to gain a fundamental understanding of the role semiosis plays in information processes. The object of the study is the structure of various types of signs, and a determination of the relationship between sign structure and information properties.

The study started with four specific goals: 1) the determination of the internal structure of various sign categories and the relationship between internal and external structure; 2) formal explication of information including its relationship to semiotic processes and a determination of the essential dimensions of semiosis; 3) a systematic listing of all information measures that have appeared in the literature, along with the known properties of each, classified according to their dimensionality and semiotic interrelationships; and 4) the development of a "kind of directory of results, methods, and key questions in this area which can orient researchers, students, practitioners to facilitate the synthesis and evaluation of research."

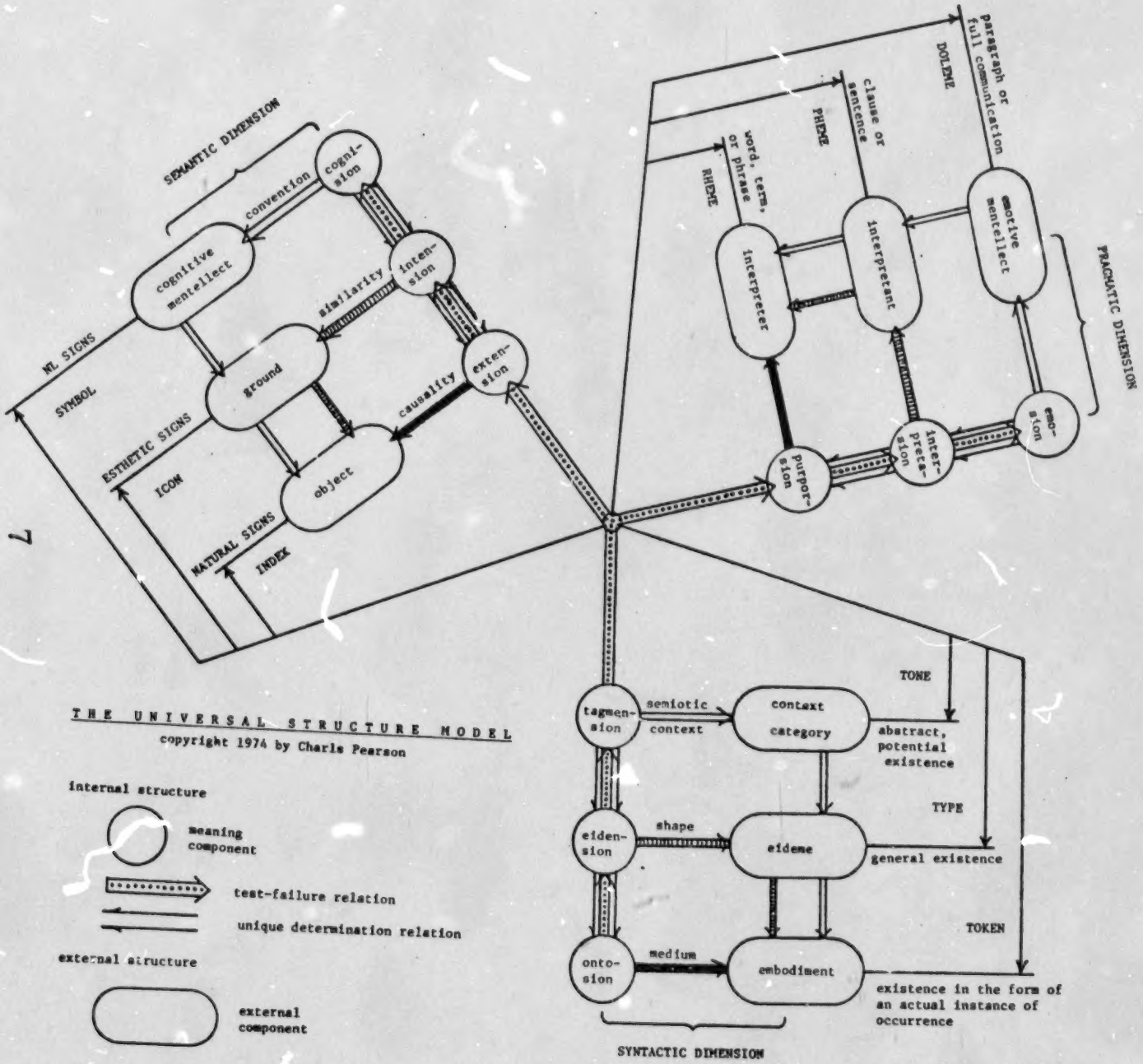
Because of the logical priority of the first two goals over the other two, research has concentrated on these. Our progress will be reported under the following six headings: 1) Theory of Sign Structure; 2) Syntactic Developments of Theory; 2) Investigations into Syntactic Shape; 4) Investigations into Semantic and Pragmatic Structure; 5) The Semiotics Lab; and 6) Other Activities.

I. THEORY OF SIGN STRUCTURE

Our purpose in developing a theory of sign structure is to have a tool for explicating the nature of information measurement and its relationship to semiotic processes and for classifying information measures according to their semiotic dimensionality and interrelationships. A theory of sign structure useful for these purposes has evolved gradually over the past three years and its broad outlines have now become quite clear. The theory is called the "Universal Structure Model," and it may be summarized by the following diagram. The theory is further elaborated in [10,11,14,15,16,17,18,19,22 & 25].

The model is universal in the sense that the structure of all categories of signs is displayed in this one model. For instance an index has only the first, or lowest, level of semantic structure which includes the object and extension of the sign; an icon has the first two levels of semantic structure which include the ground and intension of the sign in addition to the first level structure; and a symbol has all three levels of semantic structure including a cognitive mentellect and cognision.

This theory is an outgrowth of the dissertation research of one of the authors (C.P.) into the structure of the symbolic rheme reported in [14]. In that work the meaning of a sign is identified with its internal structure. A separate report on the various senses of the word 'meaning' found in a survey of 20th century literature on meaning and semiotics is being prepared for publication [15]. Another report [16] which


SEMANTIC DIMENSION
cogni-sion
convention
cognitive mentellect
intension
similarity
ground
extension
causality
object
NL SIGNS
SYMBOL
ESTHETIC SIGNS
ICON
NATURAL SIGNS
INDEX
RHEME
word, term, or phrase
PHEME
clause or sentence
DOLEME
paragraph or full communication
interpretant
interpreter
emotive mentellect
PRAGMATIC DIMENSION
purport-sion
interpre-tation
emo-sion
THE UNIVERSAL STRUCTURE MODEL
copyright 1974 by Charls Pearson
internal structure
meaning component
test-failure relation
unique determination relation
external structure
external component
tagmen-sion
semiotic context
context category
abstract, potential existence
TONE
eiden-sion
shape
eidene
general existence
TYPE
onto-sion
medium
embodiment
existence in the form of an actual instance of occurrence
TOKEN
SYNTACTIC DIMENSION

explicates the distinction between internal and external structure of signs, explicates the number of sign components, and motivates the identification of meaning with internal components is also in preparation. Some additional remarks concerning this theory were published in [19, p3-10].

II. SYNTACTIC DEVELOPMENTS OF THEORY

The universal structure model predicts three levels of syntactic structure: ontotic, eidontic, and tagmatic. The ontotic level contains the medium and ontosion of the sign and is related to the embodiment and other token characteristics. The eidentic level contains the shape and eidension of the sign and is related to the type characteristics. The tagmatic level contains the semiotic context and tagmension of the sign and is related to the context category and other tone characteristics.

In the syntactics of natural language these levels may be identified with phonetics, morphophonemics, and tagmatics, although this identification has not been explicated yet. Instead, efforts have been concentrated on using these predictions to explicate the statistical theory of syntactical communication. Progress has been quite rapid and it appears that this is the most natural explication for this theory. There is one communication component for each syntactic level. The component for the tagmatic level has been called "information source" [1, chap. 6]; the eidontic component is called "encoder," or "decoder"; and the ontotic component is the channel. The semiotic properties associated with tone, type, and token phenomena and especially such observed regularities as the rank-frequency law of

Zipf and Estoup may be used to understand the communication processes associated with each component. So-called "Information Theory" (the Shannon calculus) is seen as a theory of modal statistics which is applicable throughout the syntactic dimension (as well as to other--nonsemiotic relations), since it is the mathematical idealization of many empirical relations. This approach has been incorporated into the class notes for a senior level course on communication processes [11] and makes these processes quite easy to explain. A textbook incorporating the universal structure model approach is in preparation [22].

III. INVESTIGATIONS INTO SYNTACTIC SHAPE

Original investigations into the nature of syntactic theory have been concentrated at the eidontic level with the semiotic concept of shape being of prime interest. In many kinds of signs, shape is primarily concerned with length and pattern. In 1965 Kolmogorov proposed a measure of shape that is mainly a measure of the pattern. This has been called "algorithmic information" and "complexity" [5]. A historical and tutorial paper was prepared giving a semiotic analysis of this measure as a background for a future explication of the concept of complexity [21]. This paper has been accepted for publication in the International Journal of Computers and Information Science.

An instrument, called an 'eidontic deviometer,' or 'eidometer,' has been developed to measure the deviation of the shape of a natural language sign from a hypothetical norm, or average shape, of a sign in a given

natural language. Measurements on artificial word forms using this instrument are both reliable and precise. The instrument has been improved and rebuilt during the past year. A paper presenting this development is in preparation [20]. Miller, Bruner, and Postman [6] have shown that the interpretation of signs is affected by their shape. The eidometer should enable a precise measurement of this phenomenon and thus lead to a better understanding of the interpretation process. An elementary tachistoscope has been purchased to use with the stimuli measured by the eidometer for the purpose of refining the Miller-Bruner-Postman experiment. Success in this would lead to a direct measurement of the redundancy curve for natural language. This measurement has never been made before, although Shannon determined both upper and lower bounds for this curve [23].

From a count of 5-1/2 million letters in a corpus of standard American (the Brown Corpus) a table of polygram frequencies has been prepared. Previous tables for American English available to the public were at least half a century old and their generality is at best suspect, having most likely come from counts of military documents. Several later and more general counts are available within the CIA and the Department of Defense, but their access requires a security clearance and a 'need to know.' These tables are useful for the generation of artificial word forms and the study of redundancy in natural language. A paper presenting these tables is in preparation [13]. During the analysis of this count data, a rank-frequency regularity was discovered among the letters. However, unlike the rank-frequency law of Zipf and Estoup for words,

which is log-log in nature, the regularity for letters is log-linear in nature. A preliminary literature search shows no previous mention of this regularity. All available data for other alphabets and phonemic systems was analyzed and the relationship holds in every case. Much more analysis is required before a paper can be written, but it is hoped that we have discovered a universal relation for the shape elements of a system of discrete signs.

It should be mentioned that Garner's work [4] which is analyzed in a review which is discussed in section VII [12] is also pertinent to a study of shape in semiosis.

IV. INVESTIGATIONS INTO SEMANTIC AND PRAGMATIC STRUCTURE

Another area of original investigation which has just begun concerns the semantic and pragmatic structure of signs. Although the universal structure model stems from research into natural language, this same structure should, if it has any bearing on reality at all, show up also in other disciplines which study sign processes--disciplines such as philosophy and psychology.

A preliminary argument has been developed which shows the usefulness of the universal structure model for unraveling philosophical problems. G. E. Moore, the early twentieth century British philosopher, developed a paradox which has come to be called Moore's paradox of analysis and may be stated as follows: if the analysis of the meaning of a word has the same meaning, it is trivial; but if it has a different

meaning, then it is wrong. Moore knew well that philosophers very often make correct and non-trivial analyses, but he was never able to develop a theory of analysis which overcame his own paradox. Although other philosophers have tried with varying amounts of success, the problem has never been solved completely. The most popular approach is to say that the problem lies in the formulation of the paradox, which assumes that meaning is either a single or wholistic kind of thing which is either completely the same or else altogether different. Frege and Carnap assumed that the meaning of signs has two components, but their assumptions were for entirely different purposes. Carnap was able to delineate the character of scientific analysis fairly well with his "extension" and "intension," but he was never able to handle philosophic analysis. Moore himself said he thought philosophic analysis required something like determining the same objects by the same properties but understanding or cognizing this determination in a different way. Looking at our semantic structure, however, and realizing that cognision uniquely determines intension, which in turn uniquely determines extension, while a difference in extension ensures that two terms will have a difference in intension, which in turn ensures a difference in cognision, we may state the solution as: Scientific analysis requires an identical extension with a difference in intension, while philosophic analysis requires an identical intension with a difference in cognision.

One final area we have begun to explore concerns cognitive representation. Kintsch has reported three aspects of cognitive memory which he calls "sensory," "short term," and "long term." Bruner has

reported several modes of representation, or coding, including "enactive," "iconic," and "symbolic." He has studied the sequence in which these capabilities develop in children and the rate at which signs can be processed using the various modes of representation. It would appear as if there were just one form of coding associated with each aspect of cognitive memory. However, the experimental data is not clearcut on this question because of confounding effects.

An experimental program is being designed to critically isolate each memory aspect and the mode of representation that is associated with it. The first experiment, to isolate and determine the characteristics of iconic coding, uses an interference effect discovered by Siegmann [24]. A set of stimuli, called "iconic circles" and consisting of overlapping circular disks masked from out-of-focus photographs of abstract ink blobs, has been developed and several sets of exploratory trials have been run. The interference effect is well marked and can be detected easily. The preliminary results are now being prepared for publication [17]. New and improved stimuli are being prepared; the experiment is being redesigned to incorporate the preliminary results; and a new experiment, using children, is being developed to verify Bernbach's results and match them to ours.

The advantage of achieving an answer to this question is to allow quantitative measurements of psychology to be used in future investigations of semantic structure. For instance, memory span times, processing rates, and age of development are all quantitative measurements, and all run in the sequence: index, icon, and symbol. It would be interesting to

evaluate the semantic generalization studies of Razran and the Russian school of psychologists from this same aspect of three components of semantic structure.

V. THE SEMIOTICS LAB

A semiotics lab has been developed which supports several courses in information science. The initial development was reported in [18]. A lab manual has been written for the use of the students in this lab [8]. During the past year the lab was moved into a larger room with increased area for housing graduate students and their experimental equipment. Several pieces of equipment were added to the lab facilities including a timer for the memory coding experiment, a rudimentary T-scope for the Miller-Bruner-Postman experiment, and several eidontic deviometers. We are presently exploring means for financing the purchase of a 3-field Harvard T-scope. A corpus of technical writing from pharmacology journals, three German corpuses (including one complete novel) and a corpus of aphasic text (all in machine readable form) were acquired during the year. Efforts are now underway to translate the code of the German text tapes to B-5500 compatible code. The systems software and all the instrumentation and analysis programs for the semiotics lab have been completely redesigned and reprogrammed during the year to improve both speed and capability. For instance, the type-token analysis program which formerly would handle samples with a maximum size of only 2,000 word tokens will now accommodate samples of up to 40,000 word tokens.

The Markov word generator which formerly took 10 minutes (time-shared) to generate 50 words now takes only about 10 seconds to generate the same number of words on the same basis. The most important new analysis program added to our library this year was a non-linear least squares estimation program with extended plot capabilities.

VI. OTHER ACTIVITIES

Presentation at 12th NIR Colloquium

One of the investigators (C.P.) was invited to give a paper on the goals for long-range research in information science before the 12th National Information Retrieval Colloquium in Philadelphia [9]. We may briefly summarize the goals presented in that paper as follows: (1) Develop a theory of the structure of all categories of signs, sign systems, and sign processes; (2) Investigate the measurable properties of all sign components; (3) Explore the basic regularities existing between the measurable properties; (4) Develop theories which explain these regularities; (5) Investigate the relationship between various information processes and semiotic processes: perception, memory, recall, conception, communication, classification, recognition, decision, etc; (6) Develop lab instrumentation with which to carry out the critical experiments in the above areas.

This last is of special importance and is crucial to each of the above goals. Because the phenomena under investigation is semiotic and not physical this requires the invention of entirely new concepts in scientific instruments. It will therefore go hand-in-hand with goals (2) and (3).

Semiotic Society of America

One of the investigators (C.P.) attended the North American Semiotics Colloquium in Tampa, Florida. One of the long-term results of the North American Semiotics Colloquium was the formation of the Semiotics Society of America. The provisional officers consist of the following: Henry Hiż, University of Pennsylvania, president; Eugen Bär, Hobart and William Smith Colleges, vice-president; and Thomas Sebeok, Indiana University, secretary-treasurer. The principal investigators have been very active in the organizing of the Society, with the result that the Georgia Institute of Technology will host the Society's first annual conference in Atlanta this September 24th and 25th. One of the investigators (C.P.) was appointed director of the conference. We are also planning a semiotics lab workshop on the day before the conference. Specific conference sessions that are being organized and chaired by project staff members include: 1) Semiotic Foundations of Information Science; 2) Experimental Semiotics; 3) Mathematical Semiotics; 4) Theoretical Semiotics; 5) Medical Semiotics; and 6) Zoosemiotics--Communication between Man and Animals.

Publication Activities

W. P. Garner's latest book on information and structure, a report on ten years of research into the nature of medium and shape in the process of sign perception was reviewed for the ACM's Computing Reviews [4, 12]. The review concentrates on an analysis of the semiotic concepts involved.

A textbook which uses semiotic concepts to explicate the measurement of information is being developed [25].

One of the investigators (C.P) was appointed technical editor for semiotics on the staff of Foundations, the American Society for Information Science newsletter for Foundations of Information Science Special Interest Group. A report on software development for the semiotics lab [2], and another on the Universal Structure Model and other semiotic investigations [10] have been accepted for publication by that newsletter. Another report on progress in software development has been prepared for The Semiotic Scene, newsletter of the Semiotic Society of America [3]. One of the investigators (C.P.) gave a seminar on Peirce's Theory of Signs and the Universal Structure Model to the Interdisciplinary Program in Information Science at the University of Pittsburgh [7]. An interim progress report on this project appeared in [19].

REFERENCES

1. Ash, R. Information Theory. Interscience, 1965.

2. Flowers, J. "Software Development Progress at Georgia Tech's SemLab." To appear in Foundations, an American Society for Information Science Newsletter, 1976.

3. Flowers, J. "Progress in Semiotics Software at Georgia Tech's SemLab." Submitted to The Semiotic Scene, newsletter of the Semiotic Society of America, 1976.

4. Garner, W.P. The Processing of Information and Structure. Wiley & Sons, 1974.

5. Kolmogorov, A.N. "Three Approaches to the Quantitative Definition of Information." Problems of Information Transmission, 1(1965), p1-7.

6. Miller, G.A., Bruner, J.S., and Postman, L. "Familiarity of Letter Sequences and Tachistoscopic Identification." J. Gen. Psychol., 50(1954), p129-139.

7. Pearson, C.R. "Peirce's Theory of Signs and the Universal Structure Model." A seminar presented to the University of Pittsburgh Interdisciplinary Program in Information Science, Pittsburgh, Pennsylvania. October 21, 1975.

8. Pearson, C.R. "Lab Manual for Semiotics." School of Information and Computer Science, Georgia Institute of Technology. Atlanta, Georgia, 1975.

9. Pearson, C.R. "Goals for Long-Range Research in Information Science." An invited paper delivered to the 12th National Information Retrieval Colloquium; May 1-2, 1975; Drexel University; Philadelphia. Presented at the session on Long-Range Planning for Basic Research in Information Science.

10. Pearson, C.R. "Information Science and the Universal Structure Model." To appear in Foundations, an American Society for Information Science Newsletter, 1976.

11. Pearson, C.R. "Communication Processes": unpublished course notes for ICS-4120.

12. Pearson, C.R. "Review of Information and Structure by Garner, W.P." To appear in Computing Reviews, 1976.

13. Pearson, C.R. "A Table of Polygram Frequencies." (In preparation.)

14. Pearson, C. R. Towards an Empirical Foundation of Meaning. Ph.D. thesis; Georgia Institute of Technology, 1976. (In preparation.)

15. Pearson, C.R. "The Senses of 'Meaning.'" (In preparation.)

16. Pearson, C.R. "The Components of Signs." (In preparation.)

17. Pearson, C.R.; Siegmann, P.J.; and Shin, K.J. "An Experimental Investigation of Iconic Coding." (In preparation.)

18. Pearson, C.R.; and Slamecka, V. "Semiotic Foundations of Information Science." In Research 1973/1974: Annual Progress Report, School of Information and Computer Science, Georgia Institute of Technology; Atlanta, Georgia; p3-7.

19. Pearson, C.R.; Slamecka, V.; Zunde, P.; Flowers, J.; Betts, R.; Shin, K.J.; and Canyas, D. "Semiotic Foundations of Information Science." In Research 1974/1975: Annual Progress Report, School of Information and Computer Science, Georgia Institute of Technology; Atlanta, Georgia; p3-10.

18

20. Pearson, C.R.; Smith, L.; and Ray, J., Jr. "The Eidontic Deviometer: An Instrument for Measuring Shape Deviance in Natural Language." (In preparation.)

21. Pearson, C.R.; and Zunde, P. "The Kolmogorov Potential as a Measure of Algorithmic Information." To appear in International Journal of Computers and Information Science.

22. Pearson, C.R.; and Zunde, P. A Semiotic Approach to the Statistical Theory of Syntactic Communication. (Textbook in preparation.)

23. Shannon, C.E. "Prediction and Entropy of Printed English." Bell Syst. Tech. Jour. 30:50-64 (1951)

24. Siegmann, P.J. Personal communication. 1975.

25. Zunde, P.; and Pearson, C.R. Information Measurement. (Textbook in preparation.)